Seismic Hazard Epistemic Uncertainty in the San Francisco Bay Area and its Role in Performance-Based Assessment

Brendon A Bradley [a)]

This paper investigates epistemic uncertainty in the results of seismic hazard analyses for the San Francisco bay area and their role in the broader picture of seismic performance assessment. Using the 2002 Working Group on California Earthquake Probabilities earthquake rupture forecast, epistemic uncertainty in the seismic hazard for several different intensity measures and sites in the San Francisco bay area is investigated. Normalization of the epistemic uncertainty for various sites and intensity measures illustrates that the uncertainty magnitude can be approximately estimated as a function of the mean exceedance probability. The distribution of the epistemic uncertainty is found to be dependent on the set of alternative ground motion prediction equations used, but is frequently well approximated by the lognormal distribution. The correlation in the hazard uncertainty is observed to be a function of the separation between the two different intensity levels, and a simple predictive equation is proposed based on the data analysed. Three methods for the propagation of seismic hazard epistemic uncertainty are compared and contrasted using an example of the 30-year collapse probability of a structure. It is observed that, for this example, epistemic uncertainty in the collapse capacity is more influential than that in the seismic hazard.

INTRODUCTION

Evaluation of the seismic risk of structures and facilities is burdened by significant uncertainties. In general such uncertainties can be classed as either aleatory or epistemic. Aleatory uncertainty is due to randomness, while epistemic uncertainty is due to the lack of knowledge of the process being observed. While the separation of uncertainties as being

[a)] Department of Civil and Natural Resources Engineering, University of Canterbury, Private Bag 4800, Christchurch, NZ.

either aleatory or epistemic is not a trivial task, from a pragmatic viewpoint, the distinction should be made based on uncertainties which the analyst can and cannot reduce (Der Kiureghian and Ditlevsen 2008).

The distinctly different nature of these uncertainties implies that different methods should be employed to propagate them in uncertainty analyses. For probabilistic seismic hazard analysis (PSHA), in particular, where the goal is to determine the probability of exceeding some level of seismic intensity measure, aleatory uncertainties are considered explicitly in the computation of the seismic hazard and result in a single seismic hazard curve, while epistemic uncertainties lead to multiple hazard curves. Furthermore, aleatory uncertainties are, in theory, random such that given a long enough period of time all possible different values of the process will be realised. Epistemic uncertainties, on the other hand, represent different possibilities of a process which has only one true but currently unknown value and are commonly therefore non-ergodic (Der Kiureghian 2005, Der Kiureghian and Ditlevsen 2008).

Treatment of seismic hazard and seismic risk using frameworks such as the Pacific Earthquake Engineering Research (PEER) centre performance-based earthquake engineering (PBEE) equation (Cornell and Krawinkler 2000) can explicitly consider epistemic uncertainties in the seismic hazard at the site of a specific structure and propagate such uncertainty to other measures of seismic performance (e.g. Baker and Cornell 2008).

This paper investigates epistemic uncertainty in the results of seismic hazard analyses of the San Francisco bay area, and their role in the seismic performance assessment of structures and facilities. Using the 2002 Working Group on California Earthquake Probabilities (WGCEP02) earthquake rupture forecast (ERF) for the San Francisco bay area, the epistemic uncertainty in the seismic hazard for several different intensity measures at several sites is investigated. Based on the observed results some features of the epistemic uncertainty are characterised including: (i) uncertainty magnitude and variation with probability of exceedance; (ii) distribution of the probability of exceedance for a given intensity; (iii) correlation of the epistemic uncertainty in the exceedance probability at different intensity levels. Three methods of various complexity and input requirements for the propagation of seismic hazard epistemic uncertainty to other seismic performance measures are compared and contrasted using an example of the collapse risk of a structure.

CONSIDERATION OF EPISTEMIC UNCERTAINTIES IN PBEE

There exist many uncertainties in seismic hazard analysis, and while an exhaustive list is

beyond the scope of this work (details can be found in, for example Kramer (1996) and McGuire (2004)), it is noted that such uncertainties can be classified as either relating to the prediction of earthquake ruptures, or the characterisation of the resulting ground motions. Typical epistemic uncertainties in earthquake prediction include (WGCEP 2003): (1) time-dependent nature of characteristic ruptures; (2) magnitude-area scaling relations; (3) fault segmentation endpoints; (4) seismogenic thickness; (5) fault slip rates; (6) relative frequency of various multi-segment ruptures; (7) amount of aseismic slip; (8) magnitude-frequency distributions; and (9) off-fault seismicity, among others. Given that an earthquake rupture occurs, epistemic uncertainties relating to the earthquake-induced ground motions observed at or near the ground surface may include: (1) Non-uniformity and occurrence time of slip across the rupture surface; (2) direction of fault rupture; (3) variation in ground motion attenuation with distance; (4) effects of geologic structures, such as basin and other topographic effects; and (5) effects of surficial soils, among others.

As will be seen, the magnitude of the aforementioned uncertainties is significant, and therefore it is prudent that they are considered via the use of logic trees (Kulkarni *et al.* 1984), which is a discrete approach used in contemporary PSHA for handling epistemic uncertainties.

Remembering that seismic hazard analyses are used as an input to determine the seismic risk of structures and facilities, then epistemic uncertainties in seismic hazard will result in uncertainty in the value of seismic performance measures which are dependent on the seismic hazard. The following section investigates the characteristics of epistemic uncertainties for a variety of ground motion intensity measures (IM's) and sites in the San Francisco bay area; while the last section investigates three methods by which epistemic uncertainties in seismic hazard can be propagated to seismic risk measures.

CHARACTERISATION OF EPISTEMIC UNCERTAINTY

Characterisation of epistemic uncertainties in seismic hazard requires determination of the magnitude, distribution and correlation structure of the uncertainty. While inevitably each of these characteristics will be specific to a particular site, several important details can be obtained by considering several sites and ground motion IM's in regions where the epistemic uncertainties are relatively well researched. Herein use is made of the detailed ERF developed for the San Francisco bay area (WGCEP 2003).

THE WGCEP02 EARTHQUAKE RUPTURE FORECAST (ERF)

The Working Group on California Earthquake Probabilities (WGCEP 2003), or WGCEP02 herein, developed a time-dependent ERF for the San Francisco bay area which is arguably the most sophisticated ERF ever developed (Field *et al.* 2005). This ERF contains all epistemic uncertainties related to earthquake occurrence discussed in the previous section. The sophisticated nature of the ERF meant that it was not feasible to quantify the epistemic uncertainty by directly considering all of the end nodes of the logic tree, but to instead use a Monte Carlo procedure based on the relative weights of the various branches (WGCEP 2003). The WGCEP02 ERF has been implemented in the OpenSHA framework (Field *et al.* 2003), an open-source code for seismic hazard analysis and was used in this study. A typical seismic hazard curve for PGA in San Francisco is illustrated in Figure 1. It is important that such a rigorous example has been considered in this study as will be later discussed.

GEOGRAPHICAL LOCATIONS CONSIDERED

Four different sites were investigated in order to consider different levels of seismicity and different dominant faults. The sites were San Francisco (37.80°N,122.42°W); Stockton (37.90°N,121.25°W); Sacramento (38.52°N,121.50°W); and San Jose (37.37°N,121.93°W). The mean seismic hazard curves for the different sites are shown in Figure 2 (all sites located on soil with $V_{s(30)}$=760m/s). These four sites span a wide range of seismic hazard with 6.1% and 1.2% in 30 year exceedance values in the region of 0.1g and 0.2g in Stockton and Sacramento, and 0.5g and 0.8g in San Francisco and San Jose (although the ERF is strictly time-dependent, for the purpose of comparison these two probabilities of exceedance are approximately equivalent to 10% and 2% in 50 years using the Poisson assumption). A time span of 30 years is used herein as was adopted in the WGCEP02 study. While two of the regions each represent moderate-to-low seismicity and high seismicity, different faults will dominate the hazard at each site due to their different geographical locations.

Because interest in this study relates to epistemic uncertainties, it is noted that the seismic hazard curves presented herein are computed neglecting background seismicity which has low epistemic uncertainty (it is highly constrained by instrumental seismicity). The Gutenberg-Richter ‘tail’ (WGCEP 2003) associated with the magnitude distributions of individual faults is however considered.

GROUND MOTION PREDICTION EQUATIONS CONSIDERED

The WGCEP02 ERF, as the name implies, provides only the spatial and temporal occurrence of earthquake ruptures and not the resulting ground motions at the four considered sites. For this purpose ground motion prediction equations (GMPE's) are required. In this study two sets of GMPE's were used. The first set of GMPE's were those used in the 2002 update of the United States National Seismic Hazard Maps (Frankel *et al.* 2002), specifically: Campbell and Bozorgina (2003), Boore et al. (1997), Abrahamson and Silva (1997), and Sadigh et al. (1997). For brevity these models are referred to as CB03, BJF97, AS97, and S97, respectively. The second set of GMPE's were developed as part of the Next Generation Attenuation (NGA) project which included: Campbell and Bozorgnia (2008), Boore and Atkinson (2008), and Abrahamson and Silva (2008). This second set is referred to as CB08, BA08 and AS08, respectively. It is noted that the other two empirical GMPE's developed in the same phase of the NGA project (Chiou and Youngs 2008, Idriss 2008) were not currently available in OpenSHA at the time this study was conducted.

Figure 3 illustrates the mean seismic hazard curves obtained for one-second spectral acceleration in San Francisco using the two different sets of GMPE's and the WGCEP02 ERF. Clearly in the case of Figure 3a, the magnitude of epistemic uncertainty due to different GMPE's will be of the same order of magnitude as the epistemic uncertainty due to the prediction of earthquake occurrence (e.g. Figure 1). In the case of Figure 3b however, the difference between the hazard curves obtained using the CB08, BA08 and AS08 models is significantly less (although two of the NGA models were not considered, comparisons of all of the models indicate that the same conclusion would be drawn (Abrahamson *et al.* 2008)). In the results to follow, each of the prediction equations in the two sets of GMPE's were given equal weighting (i.e. 25% for each in the first, and 33% for each in the second).

Four different response spectral quantities, namely: PGA, S_a(0.5s), S_a(1.0s), and S_a(2.0s) (all 5% damped) were considered to investigate the variation of epistemic uncertainties with ground motion intensity measure type.

MAGNITUDE OF EPISTEMIC UNCERTAINTY

The 'magnitude' of epistemic uncertainty, as referred to herein, represents the size of the epistemic uncertainty in the ground motion hazard curves. The magnitude of epistemic uncertainty relates to the level of scientific uncertainty in the prediction of the seismic hazard, and is comprised of uncertainty in the occurrence of earthquake ruptures, and uncertainty in

their resulting ground motions (i.e. uncertainty in the ERF and ground motion prediction, GMP). Herein epistemic uncertainties in the ERF were obtained by using 50 Monte Carlo simulations (for each GMPE) of the WGCEP02 logic tree, while uncertainty in GMP was considered via the use of multiple GMPE's, each with equal weighting. The number of Monte Carlo simulations was based on a compromise between adequate representation of the epistemic uncertainty (i.e. convergence in the magnitude of the uncertainty) and computational time.

As the hazard curves for different sites, ground motion prediction equations, and ground motion intensity will all be unique then to study the trends in the epistemic uncertainty it was necessary to normalise the data so it can be presented together. For each seismic hazard curve the epistemic uncertainty was quantified using the lognormal standard deviation in the probability of exceedance, $\sigma_{\ln P(IM>im)}$, as a function of IM. The lognormal standard deviation was used because as the following section illustrates the lognormal distribution is a good approximation for the epistemic uncertainty in the hazard for the majority of the sites and intensity measures considered. For each IM value at which $\sigma_{\ln P(IM>im)}$ was computed, the mean exceedance probability, $\mu_{\ln P(IM>im)}$, was also computed thus yielding pairs of $\left(\mu_{\ln P(IM>im)}, \sigma_{\ln P(IM>im)}\right)$ data for each site and intensity measure. Figures 4a and 4b illustrate the variation in $\sigma_{\ln P(IM>im)}$ with $\mu_{\ln P(IM>im)}$ for the four different sites and ground motion IMs (i.e. there are 4 sites x 4 intensity measures = 16 lines for each GMPE in Figures 4a and 4b) when only ERF uncertainty is considered. While there is obviously scatter in the results for the different sites and IM's, it can be seen that the normalisation allows, in part, for the different seismicity and ground motion measures, and that there is a clear trend for increasing epistemic uncertainty as the exceedance probability reduces. Once the data were plotted in the form shown in Figures 4a and 4b, trends in the magnitude of epistemic uncertainty for the different sites, ground motion measures, and GMPE's were investigated. Figures 4a and 4b illustrate the deaggregation of the data based on the different GMPE's. In the case of Figure 4a there is a clear trend that ERF uncertainty gives larger uncertainty in the seismic hazard when the S97 and AS97 models are used compared to the BJF97 and CB03 models. Figure 4b does not indicate any dependence of epistemic uncertainty magnitude on the CB08, BA08, or AS08 models, for all of which the magnitude is similar to the BJF97 and CB03 models. No obvious dependence on geographical site or intensity measure type was observed when only a single GMPE was used.

Figures 4c and 4d illustrate the magnitude of epistemic uncertainty when considering uncertainty in the ERF and GMP using various GMPE's for the four sites and four ground motion IM types. Similar, to Figures 4a and 4b the magnitude of the epistemic uncertainty increases with reducing probability of exceedance. Also, the magnitude of epistemic uncertainty when considering uncertainty in both ERF and GMP is significantly larger than when considering ERF uncertainty alone. Table 1 illustrates at the 1.2% in 30 year probability of exceedance, the range of values for the dispersion in the seismic hazard based on the results in Figure 4, as well as the range of the ratio of the 84^{th} percentile to median exceedance probability for this given mean exceedance probability. As can be seen from Figures 4a, 4b and the second column of Table 1, the dispersion with only ERF uncertainty is similar when using the two different sets of ground motion prediction equations. However, when considering GMP uncertainty, there is a significant increase in the magnitude of epistemic uncertainty using the first set of GMPE's compared to the second. In particular, the deaggregation of lines based on ground motion IM type in Figure 4c reveals that the uncertainty is much lower when predicting PGA than the other three spectral quantities (while there is also evidence in Figure 4d of lower uncertainty in PGA, it is less pronounced).

Figure 5 illustrates the seismic hazard curves for two different sites and ground motion measures which are annotated in Figure 4c. It can be seen in Figure 5a that most of the seismic hazard curves from the different GMPE's are overlapping, an indication that the differences in the ground motion prediction equations are of a similar order as the uncertainty in the ERF itself. Figure 5b illustrates the case where there is significantly larger disparity in the seismic hazard curves using the different GMPE's. In particular the BJF97 model gives a significantly lower hazard than the other three models. The second set of GMPE's (i.e. CB08, BA08, AS08) give a smaller difference in seismic hazard (as evident from Figure 3), and therefore the total epistemic uncertainty using these equations is less than using the first set of GMPE's. Campbell and Bozorgnia (2007) however suggest that because of the similarity in the predictions of the NGA models due to similar theories and empirical data, additional models for epistemic uncertainty in GMP could be used to better represent the true epistemic uncertainty. Hence, the values in Figure 4d can be considered lower than the true epistemic uncertainty.

As noted by Abrahamson (2006) logic trees are commonly interpreted to represent the uncertainty in seismic hazard analysis, however in reality they represent the range of available alternative scientific models. A consequence of this is that using available models for a site

with little or no data will indicate smaller epistemic uncertainty compared with a well studied site with many available models, when clearly the poorly studied site will have a larger epistemic uncertainty (Abrahamson 2006). Hence the results shown in Figure 4, which represent the epistemic uncertainty for sites with comprehensive alternative scientific models, can be used as somewhat of a lower bound for such unstudied sites. Additionally, as current ground motion hazard maps for the San Francisco bay area provide only the mean exceedance rate or probability for a given ground motion intensity then Figure 4 also provides a means to approximately consider the magnitude of seismic hazard epistemic uncertainties in the performance-based assessment of facilities if site-specific PSHA details (i.e. logic-tree results) are not available.

DISTRIBUTION OF EPISTEMIC UNCERTAINTY

As noted earlier, each different PSHA performed via a single branch of a logic tree results in a single ground motion hazard curve which contains only aleatory uncertainty. All of the different possible combinations of the logic tree give different possible hazard curves, which represent the epistemic uncertainty. The mean hazard curve can be determined as the weighted average of the probability of exceedance values for a given IM from each of the different hazard curves. The mean hazard curve is typically that presented (e.g. Frankel *et al.* 2002, Petersen *et al.* 2008) for the design of structures, although some note that other possible options should be considered (Abrahamson and Bommer 2005, McGuire *et al.* 2005, Musson 2005). For each level of ground motion intensity, IM, the n different ground motion hazard curves from the logic tree provide n different values for the probability (or rate depending on the PSHA formulation used (Field *et al.* 2003)) of exceeding that level of IM. Thus from the n different probabilities it is possible to determine the empirical distribution of the exceedance probability of a specified IM value.

When investigating the empirical distributions for the epistemic uncertainty in the seismic hazard curves for a single ground motion prediction equation (i.e. only epistemic uncertainty in the ERF) it was found that over the range of different sites, intensity measures, and intensity measure levels the lognormal distribution frequently provided an acceptable fit based on the Kolmogorov-Smirnov goodness-of-fit test at the 95% confidence level (Ang and Tang 1975). This observation may be a result of the assumption that fault dimensions and slip rates in the ERF are typically defined as having normal or lognormal distributions (WGCEP 2003).

Figure 6a illustrates the cumulative distribution for the epistemic uncertainty when

predicting a PGA = 1.0g at the San Francisco site using the CB03, BFJ97, AS97, S97 models, while Figure 6c illustrates the probability values from the 50 Monte Carlo simulations conducted for each GMPE. While the difference in the mean probability value from each of the GMPE's range from $6x10^{-4}$ – $2x10^{-3}$, it is clear that there is significant overlap of the individual simulations from the different GMPE's, and that this results in a total epistemic uncertainty which is well approximated by the lognormal distribution (Figure 6a).

Figures 6b and 6d show the same results as Figures 6a and 6c, except for the distribution of $S_a(0.5s)$=0.5g at the Stockton site. Clearly, the BJF97 model leads to significantly lower exceedance probability values than the other three models, with the deviation becoming more apparent as the probability of exceedance reduces, and resulting in a cumulative distribution which departs significantly from the lognormal distribution. Figure 5b illustrates that using the BJF97 model for $S_a(0.5s)$ leads to lower hazard estimates over a wide range of IM values, but Figure 5a illustrates that it is in agreement for other IM types. This may lead one to suggest that the logic tree weights applied to the different GMPE's should be a function of the IM type and value considered, as advocated by Scherbaum et al. (2005). Although not shown here, when using the CB08, BA08 and AS08 models the distribution of epistemic uncertainty resembled that for Figures 6a and 6c, which is to be expected as it is apparent from Figure 3b that there is little difference between the mean hazard estimates using this set of GMPE's.

The observation that the lognormal distribution provides a good representation of the epistemic uncertainty is desirable because the lognormal distribution is completely defined by its mean and standard deviation. The mean is that which is commonly provided in literature, while the standard deviation could be estimated based on the results of the previous section (if site-specific analysis is not viable). However, it should be kept in mind that the above observations apply to the specific data examined and may vary when other tectonic regimes and GMPE's are considered.

CORRELATION

The correlation within the epistemic uncertainty at various levels of ground motion intensity, as will be seen, is important in the process of propagation of the uncertainty in the performance-based seismic risk assessment of structures and facilities. The correlation of epistemic uncertainties in the ground motion hazard relates to the dependence of the epistemic uncertainty at different levels of IM. Figure 7 illustrates the simulation of ground motion hazard curves using the lower and upper bound correlations of zero and one, respectively.

Comparison with the hazard curves presented in the remainder of this manuscript (e.g. Figure 1) illustrates that the correlation structure of ground motion hazard curves observed in practice is much closer to the perfect correlation case than being uncorrelated. In fact, Figure 7a illustrates that if care is not taken regarding the correlation coefficient it is possible to generate hazard curves which are not one-to-one (i.e. not monotonically decreasing for increasing intensity) and therefore impossible.

For a single site and ground motion IM it is possible to estimate the correlation between probabilities of exceedance at two different IM levels using the sample correlation coefficient:

$$\rho_{i,j} = \frac{\sum_{k=1}^{n}\left(\ln P(im_i)_k - \overline{\ln P(im_i)}\right)\left(\ln P(im_j)_k - \overline{\ln P(im_j)}\right)}{\sqrt{\sum_{k=1}^{n}\left(\ln P(im_i)_k - \overline{\ln P(im_i)}\right)^2 \sum_{k=1}^{n}\left(\ln P(im_j)_k - \overline{\ln P(im_j)}\right)^2}} = \sum_{k=1}^{n}\varepsilon_{i,k}\varepsilon_{j,k} \qquad (1)$$

where $\rho_{i,j}$ is the correlation between $\ln P(IM > im_i)$ and $\ln P(IM > im_j)$; $P(im_i)_k$ is the kth value of $P(IM > im_i)$; $\overline{\ln P(im_i)}$ is the sample mean of $P(im_i)$; and $\varepsilon_{i,k} = \left(\ln P(im_i)_k - \overline{\ln P(im_i)}\right) \Big/ \sqrt{\sum_{k=1}^{n}\left(\ln P(im_i)_k - \overline{\ln P(im_i)}\right)^2}$ is the 'residual' of $P(im_i)_k$. The correlation between the logarithms was used because the distribution of $P(IM > im_i)$ was shown to be approximately lognormal in the previous section.

Figure 8 illustrates two typical plots of the residuals using the hazard data for PGA at the San Francisco site. Figure 8a illustrates that for similar IM values there exists a high correlation, while Figure 8b illustrates for IM values which are significantly different the correlation is weak. This seems relatively intuitive that the dependence decreases as the separation between im_i and im_j increases. Figure 9a illustrates this trend, where for three different values of im_i the variation in the correlation is computed at multiple im_j values. A simple predictive equation for the correlation coefficient was obtained by transforming the data from the form shown in Figure 9a to one where the abscissa is normalised by the value of im_i. Figure 9b illustrates the transformed San Francisco PGA data, the mean obtained using non-parametric regression, and the parametric equation given by:

$$\rho_{i,j} = 1.0 - 0.12\left[\ln\left(im_i \Big/ im_j\right)\right]^2 \qquad 0.25 \le \frac{im_i}{im_j} \le 4.0 \qquad (2)$$

Comparisons of the empirical form of the correlation coefficient defined by Equation (2)

with data from the four different sites and four IMs was found to be similar to that shown in Figure 9b. As mentioned previously, it is possible that Monte Carlo simulation of seismic hazard curves may produce physically impossible results (e.g. Figure 7a). While this is still a theoretical possibility using Equation (2), because the correlation coefficient is approximately 1.0 for similar IM values it is extremely unlikely. For example, if $im_i = 0.1$ and $im_j = 0.2$, Equation (2) gives $\rho_{i,j} = 0.94$, and Figure 3 illustrates that the mean exceedance probabilities for these IM values are ~0.5 and ~0.05, respectively (i.e. a factor of 10 different). Using Monte Carlo simulation for this scenario gives a probability of less than $1x10^{-8}$ that $P(im_i) < P(im_j)$. Thus Monte Carlo simulation of seismic hazard curves is not a practical problem when correlations are appropriately considered.

It should be noted that when performing performance-based computations which require this correlation (discussed in the following section) the major contribution to the value of the performance measure integral occurs over a 'small' region of the integration variable (e.g. Bradley and Dhakal (2008, Figure 4)) such that typically only the correlation over the range im_i/im_j = 0.3-3.0 will be important. Because of the high correlation over this small range of the integration variable the subsequent section illustrates that there is little difference between using the correlation model of Equation (2) and the perfect correlation assumption.

PROPAGATION OF EPISTEMIC UNCERTAINTY

In practice, determination of the ground motion hazard at a site is the first step in the seismic design or assessment of some engineered facility. Typically, the ground motion hazard is used to determine the level of seismic intensity a structure will be subject to for a given probability of exceedance. Some form of seismic response analysis is then performed to determine the response of the structure to this level of ground motion and consequences (repair cost, injuries, business disruption) associated with the seismic response are explicitly or implicitly considered. The PEER PBEE framework provides a robust methodology for quantification of the seismic performance of structures, utilizing the theorem of total probability in the same fashion as that to compute seismic hazard within PSHA.

For simplicity, attention here will be given to computation of the probability of collapse in some time interval, which is given by (in continuous and discrete forms):

$$P_C = \int_{IM} P\left(C \middle| IM = im\right) \left| \frac{dP\left(IM > im\right)}{dIM} \right| dIM = \sum_{i=1}^{n} w_i P\left(C \middle| IM = im_i\right) P\left(IM = im_i\right) \quad (3)$$

where $P(C|IM=im)$ is the probability of collapse given $IM=im$; $P(IM>im)$ is the ground motion hazard for the given time interval; and w_i is an integration weighting which will depend on the numerical integration procedure used.

Equation (3) illustrates that the collapse probability is obtained by combining the collapse fragility (obtained from seismic response analyses) and the ground motion hazard. In both of these relationships epistemic uncertainties exist, which should be propagated in Equation (3) to compute the uncertainty in the collapse probability. Below three methods for propagation of these uncertainties which cover a range of complexity and accuracy are discussed.

PARAMETRIC SECOND MOMENT METHOD

Based on the discrete form of Equation (3) it is possible to determine the uncertainty in the 30-year collapse probability using the method of moments (Ang and Tang 1975, Baker and Cornell 2008). Firstly, the mean (or best-estimate) of the collapse probability is given by the expectation of Equation (3):

$$E[P_C] = E\left[\sum_{i=1}^{n} w_i P(C|IM=im_i) P(IM=im_i)\right] \tag{4}$$

Making use of the linearity of the expectation operator and noting that the epistemic uncertainty in $P(C|IM=im_i)$ and $P(IM=im_i)$ is uncorrelated one obtains:

$$E[P_C] = \sum_{i=1}^{n} w_i \mu_{P(C|IM=im_i)} \mu_{P(IM=im_i)} \tag{5}$$

where μ_Z is the mean of Z. As the discrete form of Equation (3) is of the form $\sum_{i=1}^{n} X_i Y_i$ (i.e. the summation of the product of uncorrelated random variables) then it can be shown (Ditlevsen 1981) that the variance of the collapse probability is given by:

$$Var[P_C] = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \left\{\sigma_{C_i,C_j}\sigma_{IM_i,IM_j} + \mu_{IM_i}\mu_{IM_j}\sigma_{C_i,C_j} + \mu_{C_i}\mu_{C_j}\sigma_{IM_i,IM_j}\right\} \tag{6}$$

where C_i and IM_i are shorthand notation for $P(C|IM=im_i)$ and $P(IM=im_i)$, respectively; and σ_{Z_i,Z_j} is the covariance between Z_i and Z_j.

Thus the second moment method makes it possible to compute the mean and variance in the 30-year collapse probability, with only knowledge of mean and covariance of the seismic hazard and collapse fragility (i.e. no knowledge of either distribution is needed). However,

the second moment approach, as the name implies, provides only the first two moments of the distribution of the collapse probability, and therefore the shape of the distribution must be assumed.

SEMI-PARAMETRIC MONTE CARLO APPROACH

Based on the parametric forms of the seismic hazard and collapse fragility it is possible to use Monte Carlo simulation to generate a non-parametric distribution of the collapse probability for a given time interval. The parametric form of the seismic hazard requires the definition of the mean, epistemic covariance, and epistemic distribution as a function of IM, while the parametric distribution of the collapse fragility curve is likely to be (but not restricted to) the lognormal distribution for both aleatory and epistemic uncertainty (Zareian and Krawinkler 2007). Thus, for simulation i, $P_{C,i}$ is obtained by generating realizations of the seismic hazard curve, $P(IM)_i$, and collapse fragility curve, $P(C|IM)_i$, and solving Equation (3). By repeating this process N times, a total of N $P_{C,i}$ values are obtained from which an empirical distribution of P_C can be constructed (Ang and Tang 1975). This approach has the advantage that full details on the distribution of the ground motion hazard curve may not be available for sites in the San Francisco bay area (as publications generally provide only the mean hazard curve), so one can use the mean hazard curve as given, and the covariance and distribution as investigated in this study to estimate the collapse probability distribution. Also, unlike the second moment method, this approach results in the full collapse probability distribution (i.e. the distribution shape does not need to be assumed).

NON-PARAMETRIC LOGIC TREE APPROACH

This approach follows directly from the logic-tree approach used to consider epistemic uncertainties in the seismic hazard. The consideration of the epistemic uncertainty in the parameters of the collapse fragility simply represents additional branches on the end of the seismic hazard analysis computation (with say m possible options). If the seismic hazard logic tree has n end-nodes, then there will be a total of nxm different values for the collapse probability which can be used to obtain an empirical distribution (as for the semi-parametric Monte Carlo approach above) for the collapse probability. As a result of this continuity, no information is lost by separating the two tasks (seismic hazard and seismic response estimation). This continuity however comes with the likely requirement that the seismic hazard and collapse estimation would have to be conducted for the same site-specific study; as

such logic tree details for general sites are not likely to be publicly available.

It should also be noted that epistemic uncertainties in many other variables in the performance-based problem (i.e. structural response, damage and loss) may be represented with continuous distributions rather than the discrete-nature of logic trees. This approach can easily be handled by using Monte Carlo simulation on the end nodes of the logic tree branches and the other continuous random variables, which is in fact desirable even for seismic hazard studies when there are extensive epistemic uncertainties (WGCEP 2003).

COMPARISON OF PROPAGATION METHODS

In order to compare the three different propagation methods described above consider the 30-year collapse probability of a structure located in San Francisco. The (hypothetical) structure has a fundamental period of T_1 = 1.0s and based on seismic response analyses it is determined that the collapse capacity has a lognormal distribution with mean and dispersion of S_a(T=1.0s) = 1.9g, and $\sigma_{\ln R,C} = 0.4$, respectively. Due to (epistemic) modelling uncertainties the mean collapse capacity is also uncertain with mean of 1.9g and dispersion $\sigma_{\ln U,C} = 0.4$ (see Zareian and Krawinkler (2007) and Haselton (2007) for methodological details). No uncertainty is considered in the standard deviation of the collapse capacity (although such higher moment uncertainties can easily be handled in the semi-parametric and non-parametric methods). Unless otherwise noted, Equation (2) is used for the correlation structure for the second moment and Monte Carlo approaches, as well as the lognormal assumption for the epistemic uncertainty in the seismic hazard.

Figure 10 illustrates the collapse capacity obtained when considering only epistemic uncertainty in the seismic hazard (i.e. $\sigma_{\ln U,C}$ is zero) using the second-moment method with various assumptions on the correlation structure of epistemic uncertainties in the seismic hazard, and assuming that the collapse probability has a lognormal distribution. Via Equation (5) it can be seen that the correlation does not affect the expected value of the 30 year collapse probability, however the covariance terms for the seismic hazard σ_{IM_i,IM_j}, in Equation (6) depend on the correlation coefficient and hence give the differing results shown in Figure 10. It can be seen that the effect of the correlation is significant with dispersion values of 0.41, 0.77 and 0.84 for none, empirical, and perfect correlation assumptions, respectively.

Figure 11a illustrates the distribution of the collapse probability (uncertainty in the seismic hazard only) using the three different propagation methods. As the non-parametric logic tree

approach makes no assumptions about the nature of the epistemic uncertainty in the seismic hazard then it can be considered the 'exact' approach. It can be seen that the three methods provide good agreement in the central portion of the distribution with some differences near the tails. This agreement between the cumulative collapse probability distribution is however for a case in which the seismic hazard epistemic uncertainty was well represented by the lognormal distribution, in situations where this is not the case it is unlikely that the parametric and semi-parametric approaches will produce as similar a result as the non-parametric Logic Tree approach.

Figure 11b illustrates the distribution of collapse probability when considering epistemic uncertainty in both seismic hazard and collapse capacity. Again the different methods provide similar results (the second moment and Monte Carlo methods are similar, and thus only one is shown), but more importantly it can be seen that the uncertainty in the collapse probability has been significantly increased when collapse fragility uncertainty is considered. This is in agreement with analytical solutions, which show that the uncertainty in the collapse capacity is $\sigma_{\ln P_C} = \sqrt{k^2 \sigma^2_{\ln U,C} + \sigma^2_{\ln U,H}}$, where $\sigma^2_{\ln U,C}$ and $\sigma^2_{\ln U,H}$ are the epistemic uncertainty in the collapse capacity and hazard respectively, and k is the log-log slope of the seismic hazard curve (which increases with reducing exceedance probability) (Bradley and Dhakal 2008). Thus, while it is well acknowledged that there exist large epistemic uncertainties in seismic hazard curves, the current lack of knowledge (i.e. epistemic uncertainty) in collapse prediction appears to be more significant when considering the 30-year collapse probability for the site and structure considered.

Table 2 summarises the pros and cons of the three methods of epistemic uncertainty propagation discussed in this section. While the different methods provide different accuracy, it is most likely to be the input requirements which determine which method is employed (e.g. logic tree details of the seismic hazard must be available to use the non-parametric approach).

CONCLUSIONS

This paper has investigated the character of epistemic uncertainty in the results of seismic hazard analyses for various intensity measures and sites in the San Francisco bay area, and their propagation in the seismic performance assessment of structures and facilities. The 2002 Working Group on California Earthquake Probabilities earthquake rupture forecast as well as two sets of ground motion prediction equations were used to rigorously capture the epistemic

uncertainty in the seismic hazard for the different intensity measures and sites investigated. The magnitude of the epistemic uncertainty was observed to be significant and increased with reducing probability of exceedance. With appropriate normalisation it was illustrated that the magnitude of the epistemic uncertainty can be approximately estimated as a function of the mean probability of exceedance. For the specific sites and intensity measures considered the similarity in the ground motion prediction equations was observed to have a strong influence on the magnitude of the epistemic uncertainty.

In the performance-based assessment of structures and facilities it maybe necessary to have knowledge of the distribution and correlation structure of the epistemic uncertainty. In situations in which the epistemic uncertainty in ground motion prediction was the same order as the epistemic uncertainty in the earthquake rupture forecast, the total seismic hazard epistemic uncertainty was well approximated by the lognormal distribution. When the epistemic uncertainty in ground motion prediction was notably larger than the epistemic uncertainty in the earthquake rupture forecast, the epistemic uncertainty in the seismic hazard was generally not lognormally distributed. The correlation between the epistemic uncertainty in the probability of exceedance at two different intensities was observed to be a function of the ratio of the two intensity levels and a simple equation was proposed to predict this correlation.

Propagation of seismic hazard epistemic uncertainty to estimate the epistemic uncertainty in seismic performance assessment can be addressed in several ways of varying complexity and input requirements. Comparisons between three methods for the 30-year collapse probability of a hypothetical structure illustrated that the methods yielded similar results. The epistemic uncertainty in the collapse probability of the hypothetical structure was observed the dominant contributor to the uncertainty in 30-year collapse probability compared to epistemic uncertainty in seismic hazard curves.

The observations made regarding the magnitude, distribution and correlation of the epistemic uncertainties in seismic hazard analyses are specific to the sites and intensity measures considered and may not be appropriate for use in other tectonic regions or for other intensity measures.

ACKNOWLEDGEMENTS

Financial support from the New Zealand Tertiary Education Commission is appreciated. Dr Ned Field, Mr Kevin Milner, and Dr Christine Goulet are greatly thanked for their roles in

developing the OpenSHA applications used in this study. Comments from two anonymous reviewers are also gratefully acknowledged.

REFERENCES:

Abrahamson, N. A., 2006. Seismic hazard assessment: problems with current practice and future developments. *1st European Conference on Earthquake Engineering and Seismology*, Geneva, Switzerland

Abrahamson, N. A., Atkinson, G. M., Boore, D. M., Bozorgnia, Y., Campbell, K. W., Chiou, B., Idriss, I. M., Silva, W. J. and Youngs, R. R., 2008. Comparisons of the NGA Ground-Motion Relations. *Earthquake Spectra* **24**, 45-66.

Abrahamson, N. A. and Bommer, J. J., 2005. Probability and uncertainty in seismic hazard analysis. *Earthquake Spectra* **21**, 603-607.

Abrahamson, N. A. and Silva, W. J., 1997. Empirical response spectral attenuation relations for shallow crustal earthquakes. *Seismol. Res. Lett.* **68**, 94–126.

Abrahamson, N. A. and Silva, W. J., 2008. Summary of the Abrahamson & Silva NGA ground motion relations. *Earthquake Spectra* **24**, 67-97.

Ang, A. H. S. and Tang, W. H., 1975. *Probability Concepts in Engineering Planning and Design* John Wiley & Sons, Inc., 406pp.

Baker, J. W. and Cornell, C. A., 2008. Uncertainty propagation in probabilistic seismic loss estimation. *Structural Safety* **30**, 236-252.

Boore, D. M. and Atkinson, G. M., 2008. Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01s and 10.0s. *Earthquake Spectra* **24**, 99-138.

Boore, D. M., Joyner, W. B. and Fumal, T. E., 1997. Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: A summary of recent work. *Seismol. Res. Lett.* **68**, 128-153.

Bradley, B. A. and Dhakal, R. P., 2008. Error estimation of closed-form solution for annual rate of structural collapse. *Earthquake Eng. Struct. Dyn.* **37**, 1721-1737.

Campbell, K. W. and Bozorgnia, Y., 2003. Updated near-source ground motion (attenuation) relations for the horizontal and vertical components of peak ground acceleration and acceleration response spectra. *Bull. Seismol. Soc. Am.* **93**, 314-331.

Campbell, K. W. and Bozorgnia, Y., 2007. *Campbell-Bozorgnia NGA ground motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters*.PEER Report No. 2007/02, University of California, Berkeley.

Campbell, K. W. and Bozorgnia, Y., 2008. Campbell-Bozorgnia NGA horizontal ground motion model for PGA, PGV, PGD and 5% damped linear elastic response spectra. *Earthquake Spectra* **24**, 139-171.

Chiou, B. S. J. and Youngs, R. R., 2008. An NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra. *Earthquake Spectra* **24**, 173-215.

Cornell, C. A. and Krawinkler, H., 2000. Progress and challenges in seismic performance assessment. *PEER Center News* **3**.

Der Kiureghian, A., 2005. Non-erogodicity and PEER's framework formula. *Earthquake Eng. Struct. Dyn.* **34**, 1643-1652.

Der Kiureghian, A. and Ditlevsen, O., 2008. Aleatory or epistemic? Does it matter? *Structural Safety* **31**, 105-112.

Ditlevsen, O., 1981. *Uncertainty Modeling: with Applications to Multidimensional Civil Engineering Systems*, McGraw-Hill International Book Co., New York; London, 412pp.

Field, E. H., Gupta, N., Gupta, V., Blanpied, M., Maechling, P. and Jordan, T. H., 2005. Hazard calculations for the WGCEP-2002 forecast using OpenSHA and distributed object technologies. *Seismol. Res. Lett.* **76**, 161-167.

Field, E. H., Jordan, T. H. and Cornell, C. A., 2003. OpenSHA: A Developing Community-Modelling Environment for Seismic Hazard Analysis. *Seismol. Res. Lett.* **74**, 406-419.

Frankel, A. D., Petersen, M. D., Mueller, C. S., Haller, K. M., Wheeler, R. L., Leyendecker, E. V., Wesson, R. L., Harmsen, S. C., Cramer, C. H., Perkins, D. M. and Rukstales, K. S., 2002. *Documentation for the 2002 Update of the National Seismic Hazard Maps*.Open-File Report 02-420., United States Geological Survey

Haselton, C. B., 2007. Assessing Collapse Safety of Modern Reinforced Concrete Moment Frame Buidlings., Stanford University, Stanford, CA, 312pp.

Idriss, I. M., 2008. An NGA Empirical Model for Estimating the Horizontal Spectral Values Generated By Shallow Crustal Earthquakes. *Earthquake Spectra* **24**, 217-242.

Kramer, S. L., 1996. *Geotechnical Earthquake Engineering*, Prentice-Hall, Upper Saddle River, NJ., 653pp.

Kulkarni, R. B., Youngs, R. R. and Coppersmith, K. J., 1984. Assessment of confidence intervals for results of seismic hazard analysis. *8th World Conference on Earthquake Engineering,*, San Francisco, CA.

McGuire, R. K., 2004. *Seismic Hazard and Risk Analysis*, Earthquake Engineering Research Institute, 221pp.

McGuire, R. K., Cornell, C. A. and Toro, G. R., 2005. The Case for Mean Seismic Hazard. *Earthquake Spectra* **21**, 879-886.

Musson, R. M. W., 2005. Against Fractiles. *Earthquake Spectra* **21**, 887-891.

Petersen, M. D., Frankel, A. D., Harmsen, S. C., Mueller, C. S., Haller, K. M., Wheeler, R. L., Wesson, R. L., Zeng, Y., Boyd, O. S., Perkins, D. M., Luco, N., Field, E. H., Wills, C. J. and

Rukstales, K. S., 2008. *Documentation for the 2008 Update of the United States National Seismic Hazard Maps*.Open-File Report 2008–1128, United States Geological Survey (USGS).

Sadigh, K., Chang, C. Y., Egan, J. A., Makdisi, F. and Youngs, R. R., 1997. Attenuation relationships for shallow crustal earthquakes based on California strong motion data. *Seismol. Res. Lett.* **68**, 180-189.

Scherbaum, F., Bommer, J. J., Bungum, H., Cotton, F. and Abrahamson, N. A., 2005. Composite Ground-Motion Models and Logic Trees: Methodology, Sensitivities, and Uncertainties. *Bull. Seismol. Soc. Am.* **95**, 1575-1593.

WGCEP, 2003. *Earthquake Probabilities in the San Francisco Bay Region: 2002–2031*.Working Group on California Earthquake Probabilities, USGS Open-File Report 03-214.

Zareian, F. and Krawinkler, H., 2007. Assessment of probability of collapse and design for collapse safety. *Earthquake Eng. Struct. Dyn.* **36**, 1901-1914.

Table 1: Dispersion of epistemic uncertainty at 1.2% in 30 years probability of exceedance

Ground motion equations	Epistemic uncertainty in ERF only	Epistemic uncertainty in ERF and GMP
CB03, BJF97, AS97, S97	0.3-0.6 (1.35-1.82)*	0.5-1.5 (1.65-4.50)
CB08, BA08, AS08	0.35-0.5 (1.42-1.65)	0.5-0.8 (1.65-2.20)

*values in brackets give the ratio of the 84^{th} percentile to median hazard (=exp(σ))

Table 2: summary of uncertainty propagation methods

Method	Pros	Cons
Second moment (Parametric)	No distribution shape needed for seismic hazard; Computationally efficient.	Requires assumed distribution of performance measure being calculated based on first two moments. Difficulties in handling epistemic uncertainties in higher moments
Monte carlo (Semi-Parametric)	Can consider distribution shape in epistemic uncertainty. Can be used when logic tree details not available.	Computationally more expensive than Second moment method.
Logic tree (Non-Parametric)	Allows direct consideration of the non-parametric form of epistemic uncertainty in seismic hazard.	Requires seismic hazard results from each branch of the logic tree which are not always available.

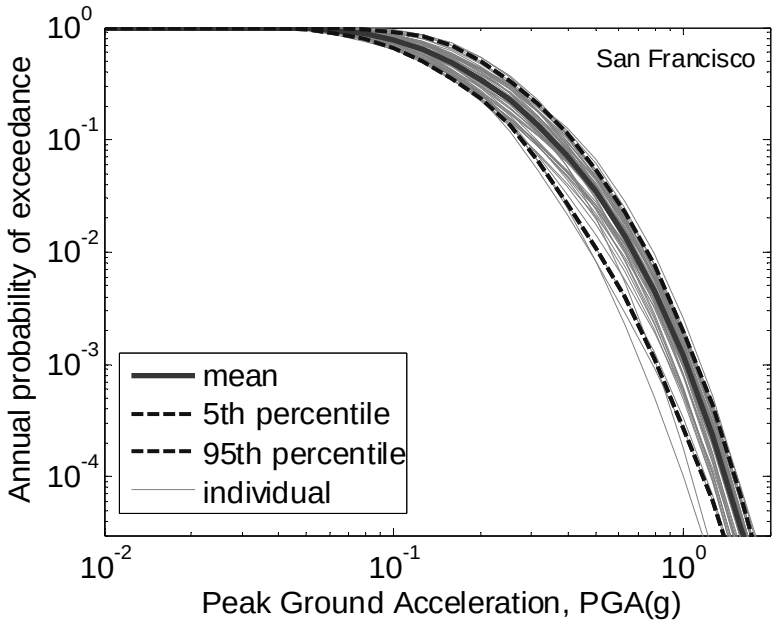

Figure 1. Illustration of earthquake rupture forecast (ERF) epistemic uncertainty in the peak ground acceleration hazard curve for a site ($V_{s(30)}$=760 m/s) in the San Francisco bay area using the Campbell and Bozorgnia (2003) prediction equation.

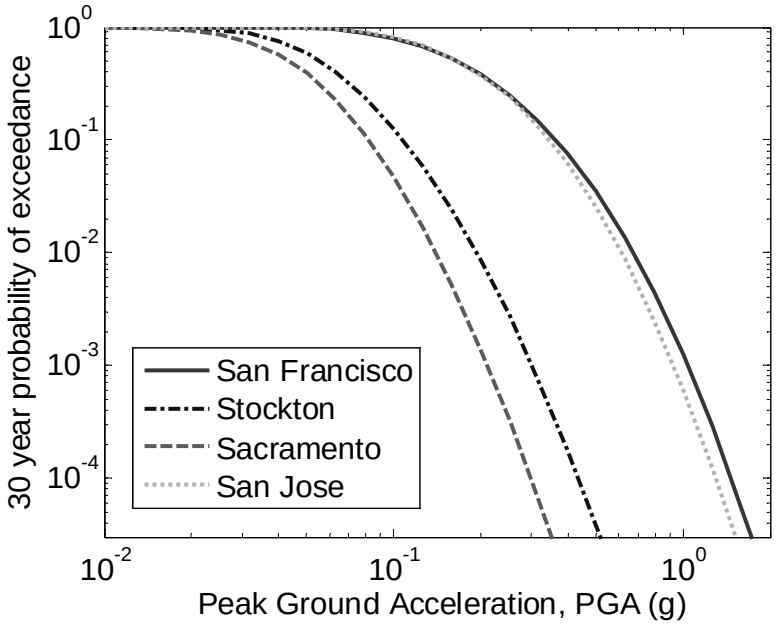

Figure 2. Mean hazard curves of the four different sites considered.

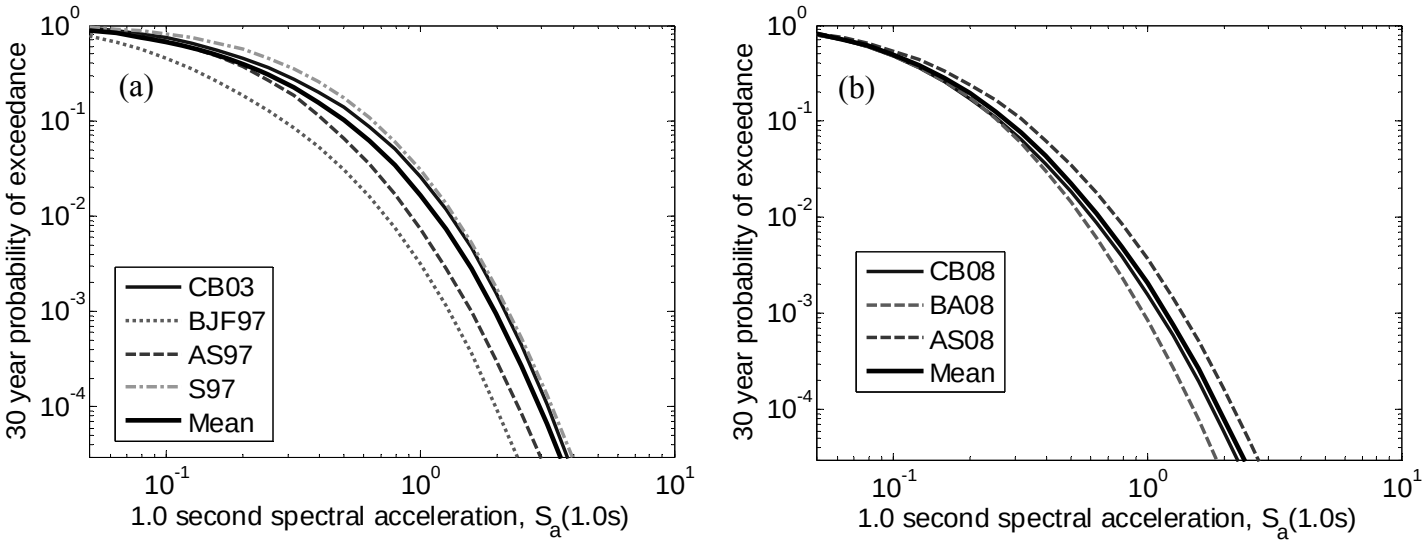

Figure 3. Illustration of the effect of various attenuation relations on the mean hazard curve in San Francisco using the two different ground motion prediction equation sets.

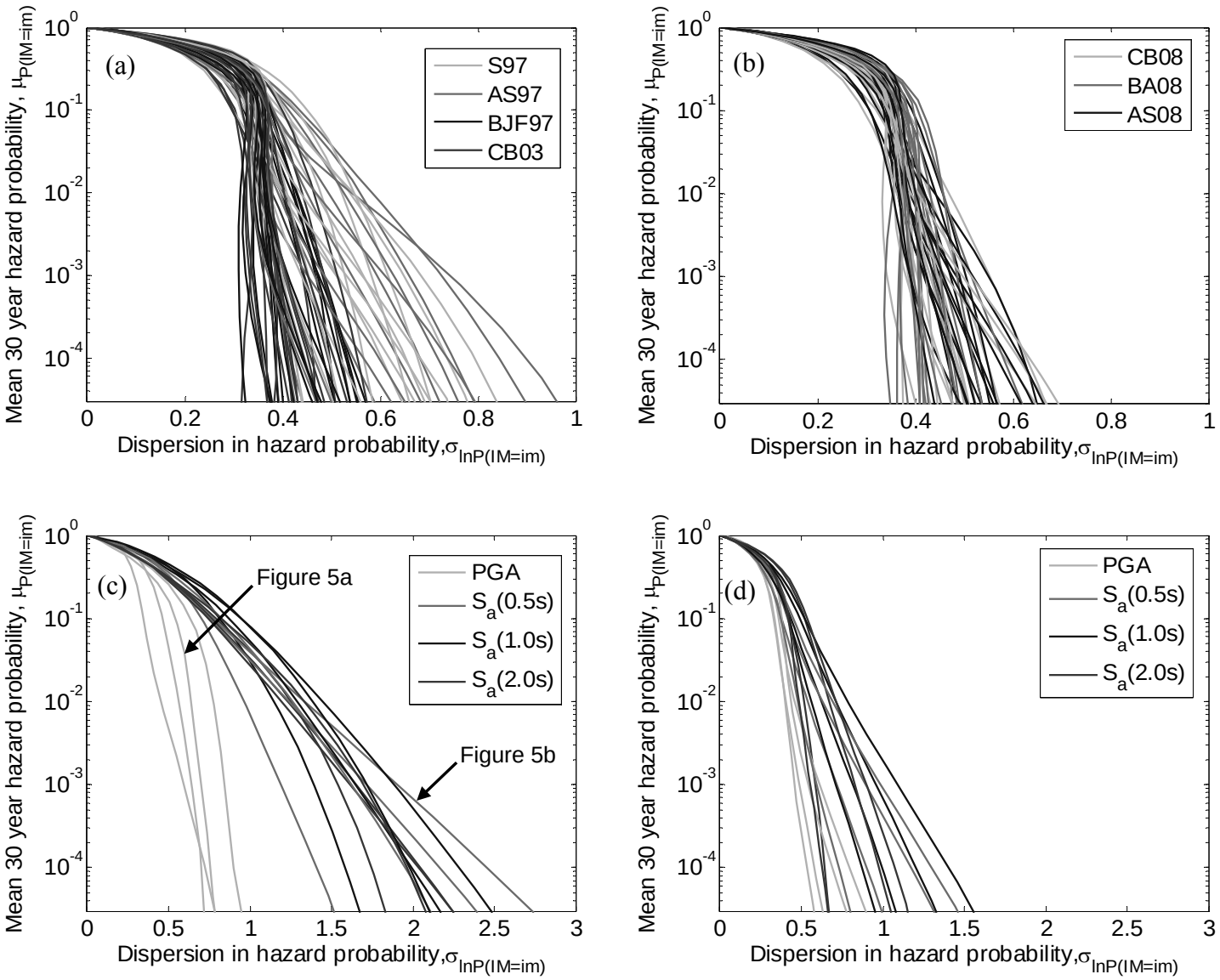

Figure 4. Magnitude of epistemic uncertainty in ground motion hazard estimates using: (a)&(b) only single ground motion prediction equation; and (c)&(d) using multiple ground motion prediction equations

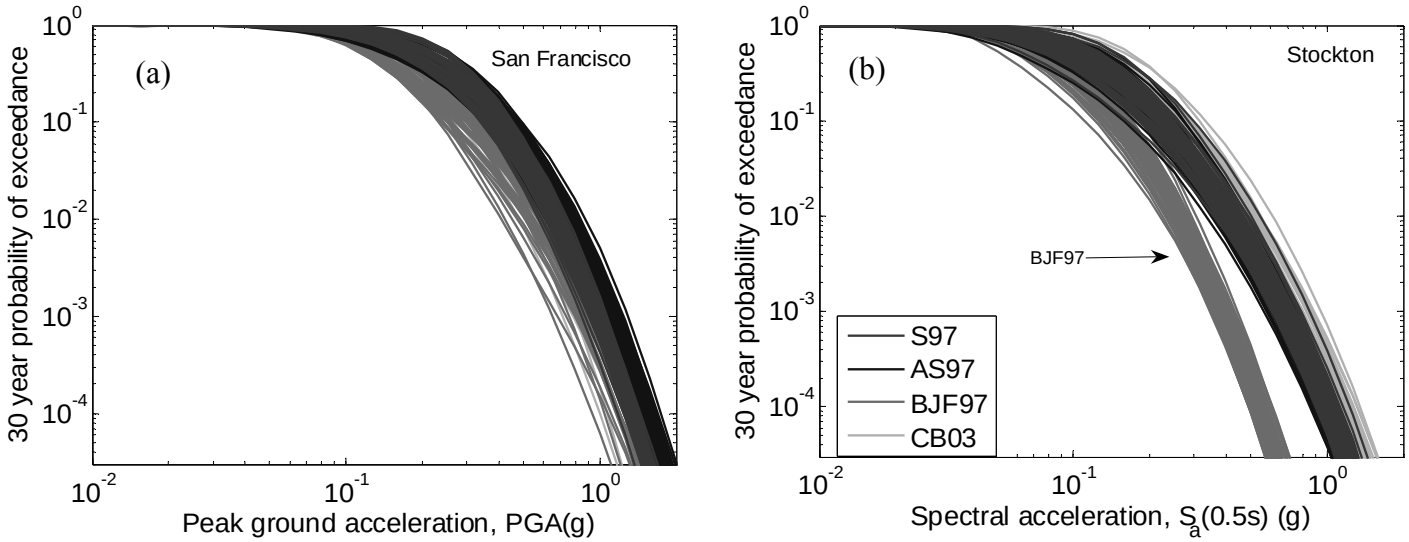

Figure 5. Example of ground motion hazard curves using different ground motion prediction equations: (a) where 'inconsistency' does not occur; and (b) where 'inconsistency' does occur.

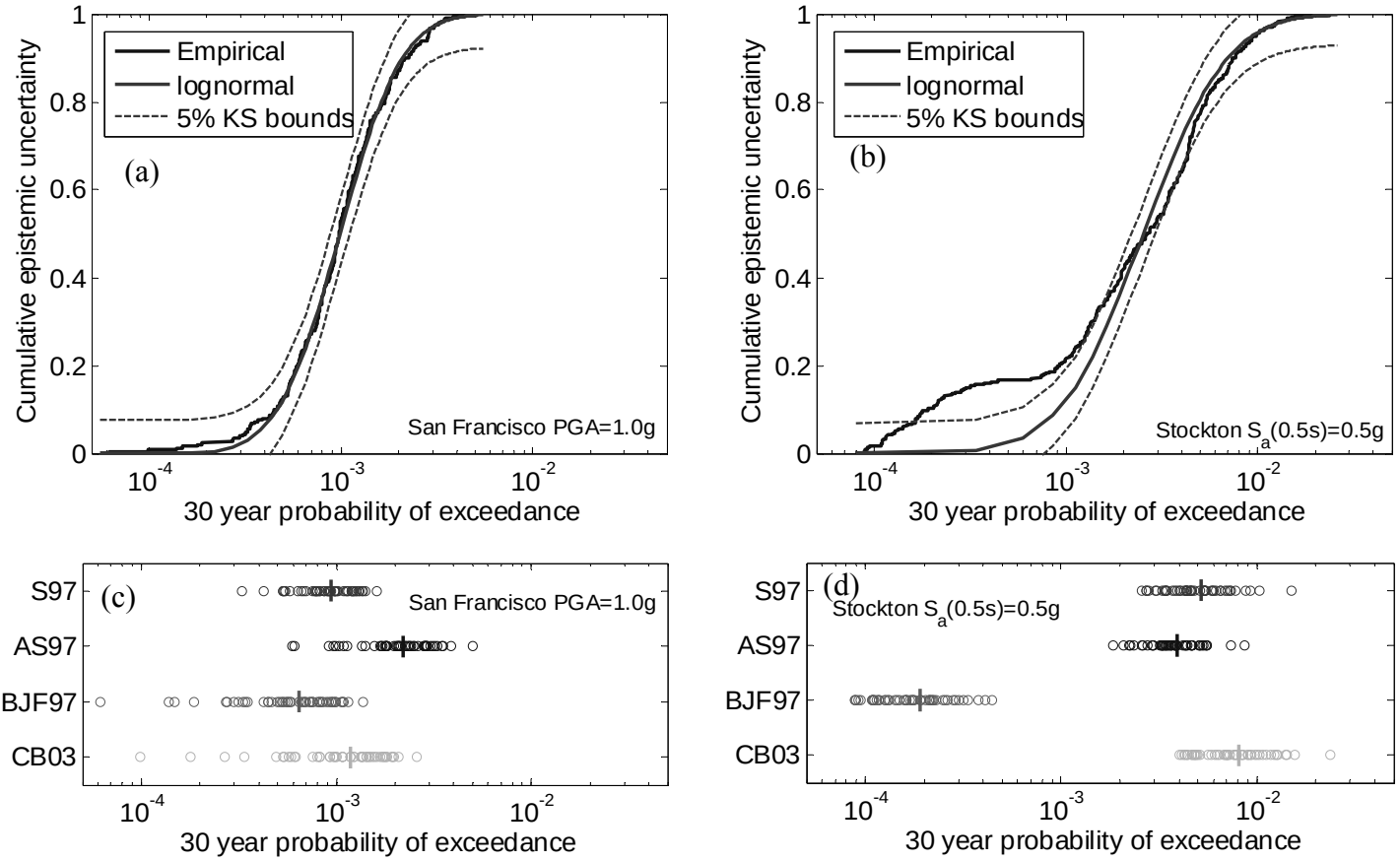

Figure 6. Distribution of epistemic uncertainty in seismic hazard curves: (a)&(b) Kolmogorov-Smirnov goodness-of-fit tests for the lognormal distribution; and (c)&(d) distribution of simulation data from different ground motion prediction equations.

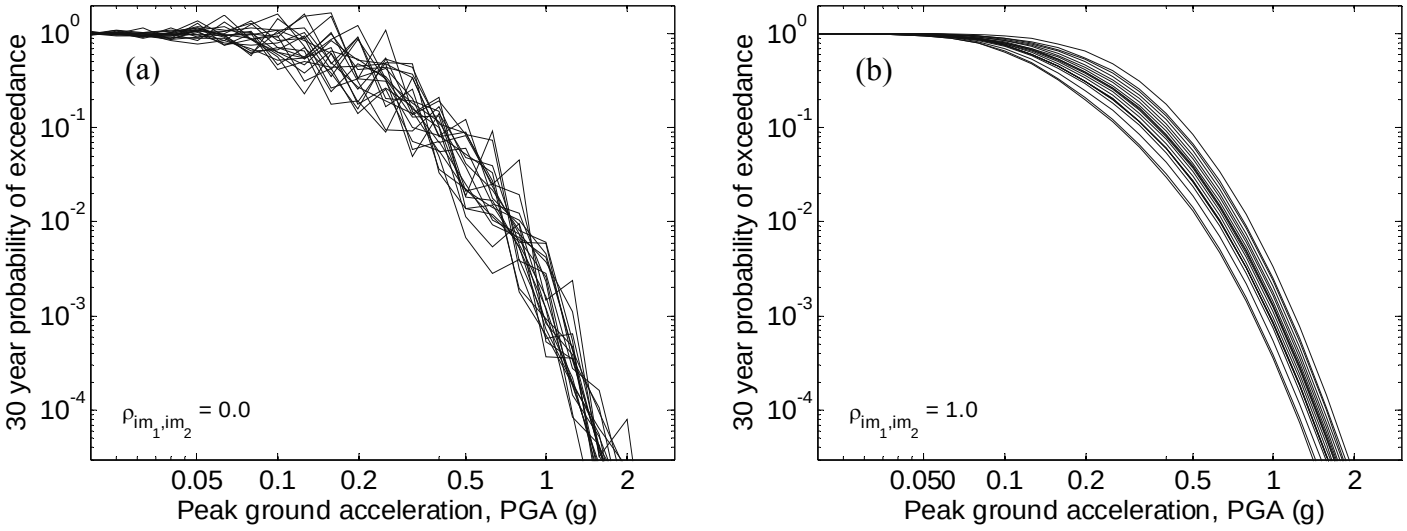

Figure 7. Effect of correlation on ground motion hazard generation

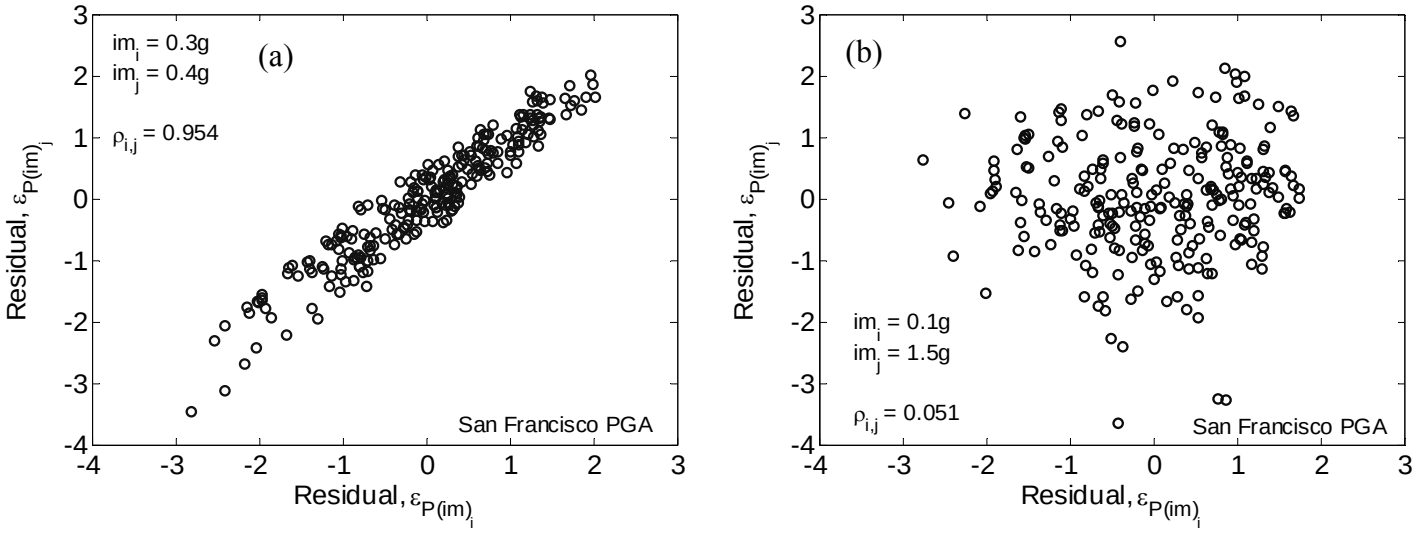

Figure 8. Illustration of correlation of epistemic uncertainty in the ground motion hazard curve for (a) intensity measure values close in absolute magnitude; and (b) intensity measure values distant in absolute magnitude.

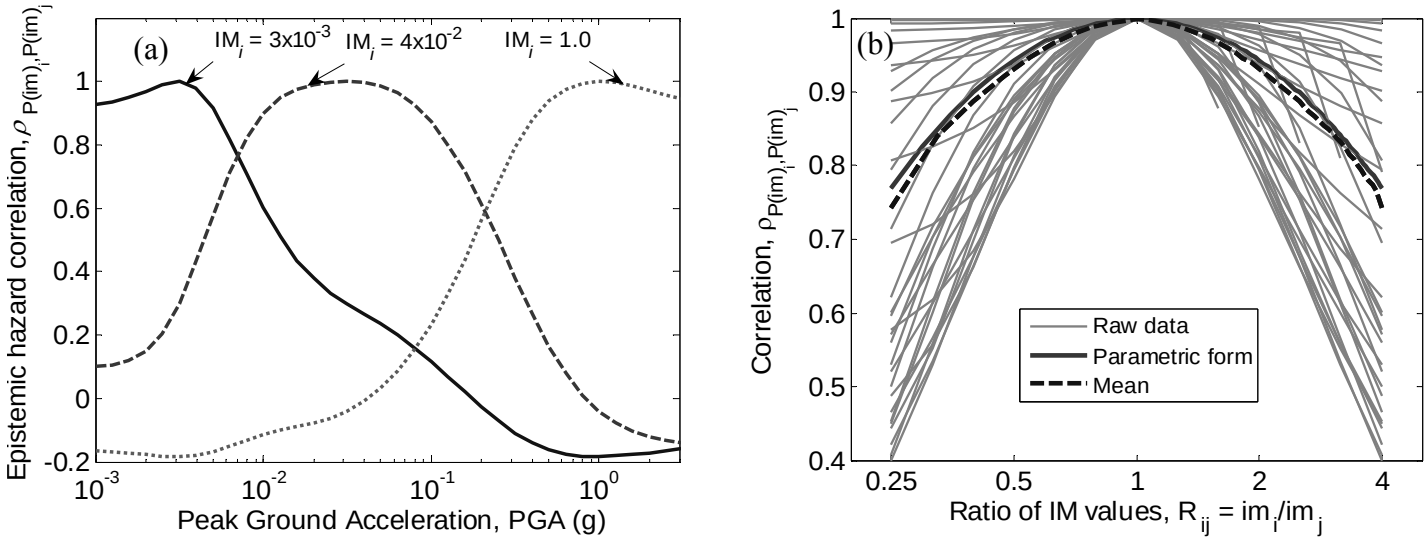

Figure 9. Correlation of epistemic uncertainty between different intensity measure values: (a) for three different intensity measure values; and (b) for all intensity measure values after normalization.

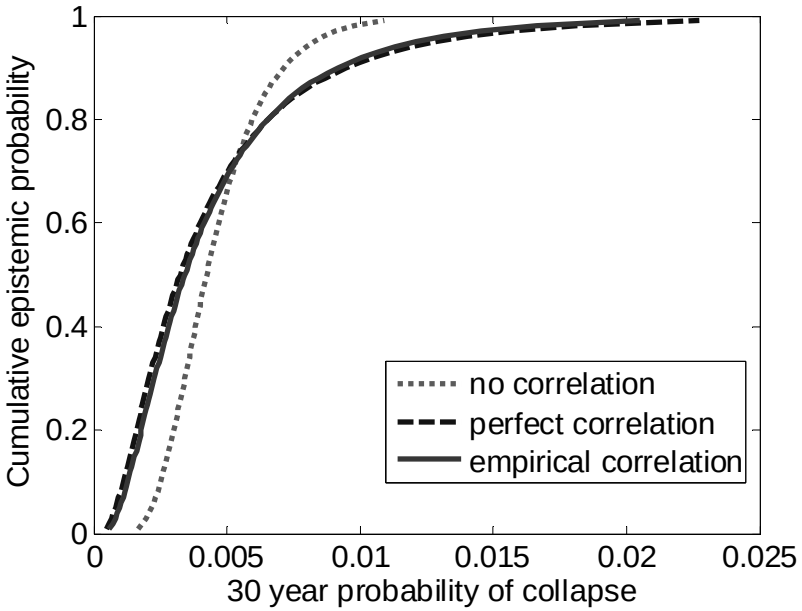

Figure 10. Effect of correlation assumption on the distribution of the 30 year probability of collapse.

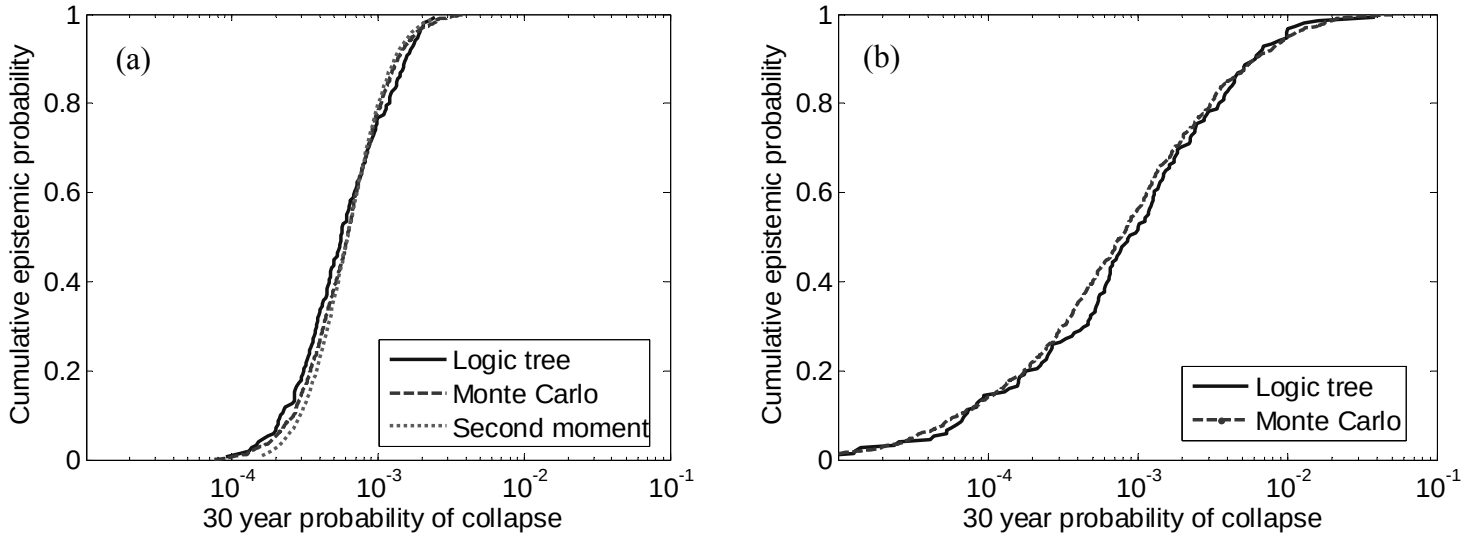

Figure 11. Distribution of collapse probability obtained using different uncertainty propagation methods: (a) only seismic hazard epistemic uncertainty; and (b) epistemic uncertainty in both seismic hazard and collapse capacity.